



Analysis of Short Time Series: Correcting for Autocorrelation

Author(s): James R. Bence

Reviewed work(s):

Source: *Ecology*, Vol. 76, No. 2 (Mar., 1995), pp. 628-639

Published by: [Ecological Society of America](#)

Stable URL: <http://www.jstor.org/stable/1941218>

Accessed: 06/12/2012 04:58

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Ecological Society of America is collaborating with JSTOR to digitize, preserve and extend access to *Ecology*.

<http://www.jstor.org>

ANALYSIS OF SHORT TIME SERIES: CORRECTING FOR AUTOCORRELATION¹

JAMES R. BENCE²

National Marine Fisheries Service, Southwest Fisheries Science Center, Tiburon, California 94920 USA

Abstract. Short time series are common in environmental and ecological studies. For sample sizes of 10 to 50, I examined the performance of methods for adjusting confidence intervals of the mean and parameters of a linear regression for autocorrelation. Similar analyses are common in econometric studies, and serious concerns have been raised about the adequacy of the common adjustment approaches, especially for estimating the slope of a linear regression when the explanatory variable has a time trend. Use of a bias-corrected estimate of the autocorrelation, either in an adjusted *t* test or in a two-stage approach, outperformed other methods, including maximum likelihood and bootstrap estimators, in terms of confidence interval coverage. The bias correction was, however, sometimes awkward to apply. It was generally better to test for autocorrelation at the 0.5 level and use ordinary least squares if the test was not significant, although this pretesting mainly helped for weak autocorrelation and small sample sizes. For the best methods, the coverage was sometimes still substantially less than the stated 95% when autocorrelation was strong, even for sample sizes as large as 50. This was true for estimates of the mean, the regression intercept, and, when the explanatory variable had a time trend, the slope. Simulation results and an example show that different adjustment methods can produce substantially different estimates and confidence intervals. Cautious interpretation of confidence intervals and hypothesis tests is recommended.

Key words: autocorrelation; autoregressive; confidence intervals; independence; regression; serial correlation; statistics; time-series; two-stage estimator.

INTRODUCTION

Data consisting of a sequence of observations collected over time are common and often unavoidable in environmental and ecological studies. It is well known that when data are collected sequentially in time the usual assumption of independence of errors is not guaranteed. The uncritical treatment of such data, as though they were a random sample, has been termed “pseudoreplication in time” (Hurlbert 1984). Often autocorrelation is positive, so that errors close in time are similar. The effect of ignoring positive autocorrelation is (1) to produce nominal confidence intervals about parameter estimates that are smaller in size than they should be, or in a hypothesis testing context to make too many Type I errors (e.g., Cochran and Orcutt 1949, Hurlbert 1984); and (2) potentially to produce less efficient estimates of parameters than could be obtained if the autocorrelation were taken into account (Cochran and Orcutt 1949). A reasonable approach, especially when dealing with the relatively short time series of most ecological studies, is to estimate the extent of first-order autocorrelation and to adjust estimates and hypothesis tests for the estimated autocor-

relation (e.g., Stewart-Oaten 1987, Carpenter et al. 1989). This approach has been widely applied in econometrics, where similar relatively short time series are common (e.g., Cochran and Orcutt 1949, King and Giles 1984, Doran et al. 1992), and concerns have been raised about the adequacy of the standard approaches.

Here I review several of the more commonly used methods that have been suggested for making this adjustment, and explore the performance of these methods, and variants of them, for small to moderate sample sizes by Monte Carlo simulation. I consider the special but important cases of estimating the mean and the parameters of a linear regression with one explanatory variable.

Estimation of the mean from relatively short time series is a common goal in environmental assessment work and in the analysis of results from unreplicated ecosystem-level experiments. In these cases, the mean being estimated is often for the differences between a “control” and “impact” site (e.g., Millard et al. 1985, Stewart-Oaten et al. 1986, Stewart-Oaten 1987, Carpenter 1989, 1990, Carpenter et al. 1989, Schroeter et al. 1993, Osenberg et al. 1994, Bence et al., *in press*), or candidate sites being considered as locations for a development or mitigation project (e.g., California Coastal Commission 1991, EPA 1993). Ecological examples of the need to estimate regression parameters from short time series abound. They range from regressing abundance or some other variable against time to detect temporal trends (Gerodette 1987, Jasby and

¹ Manuscript received 19 January 1993; revised 2 June 1994; accepted 15 June 1994; final version received 8 July 1994.

² Present address: Department of Fisheries and Wildlife, Partnership for Ecosystem Research and Management, Michigan State University, East Lansing, Michigan 48824-1222 USA.

Powell 1990, Link and Hatfield 1990), to regressing estimates of year-class strength or other measures of performance of marine fishes and invertebrates against environmental variables (e.g., Mearns et al. 1980, Botsford et al. 1989, Cury and Roy 1989). After presenting the simulation results I apply several of the methods to two example data sets.

A key result is that in many situations all of the adjustment methods undercorrect for autocorrelation by producing confidence intervals with substantially less than their nominal coverage. In the econometrics literature this result has been well established for the slope of a linear regression when the explanatory variable has a time trend. The results presented here emphasize that this problem can be quite severe for estimating the mean, or for the intercept of a regression, whether or not the explanatory variable has a trend. I consider approaches that might improve on the conventional methods, and end with a discussion of some promising alternatives.

BACKGROUND AND ESTIMATORS

The data consist of an ordered sequence, $y_i = \mu_i + \epsilon_i$, $i = 1, 2, \dots, T$, where T is the total number of times observations were collected, μ_i is the expected value at time i , and the ϵ_i are stochastic errors. With no explanatory variables, $\mu_i = \mu$, and for a linear regression of Y on X , $\mu_i = \mu + \beta x_i$. For many situations, a reasonable model would assume positive autocorrelation that decreases steadily as the time between observations increases. An autoregressive process of order one [AR(1)] is the simplest model that can produce this pattern. In this case $\epsilon_i = \rho \epsilon_{i-1} + a_i$, where the a_i are independent errors with mean zero and variance σ_a^2 , and the autocorrelation coefficient ρ is restricted to the interval $\{-1, 1\}$, with values greater than zero producing positive autocorrelation. For this model the correlation between observations k time steps apart is ρ^k .

Let $\hat{\mu}_i$ be an estimate of μ_i . The i^{th} estimated residual is $\hat{r}_i = y_i - \hat{\mu}_i$. All but one of the methods I describe below for adjusting for autocorrelation use an estimate of the autocorrelation coefficient calculated from these residuals. The widely used Prais-Winsten estimate of the autocorrelation coefficient is

$$\hat{\rho} = \frac{\sum_{i=2}^T \hat{r}_i \hat{r}_{i-1}}{\sum_{i=2}^T \hat{r}_i^2}. \quad (1)$$

Initially I used this as an estimator of the autocorrelation coefficient, except for the maximum likelihood approach. A number of variants of this estimator exist differing in the way the end points of the series are included or excluded from the denominator, and whether the estimator is adjusted when the number of components in the numerator differs from the number in

the denominator. For moderately small samples (10–50) and positive autocorrelation, the Prais-Winsten estimator is less negatively biased than other conventional competitors. An alternative symmetric estimator (Dickey et al. 1984) was also tried, with less satisfactory results. In some experiments I replaced the estimator for ρ in Eq. 1 by a bias-corrected estimator suggested by Doran et al. (1992). The bias-corrected estimator is

$$\hat{\rho}^* = \hat{\rho} + \frac{(m+1)(1+\hat{\rho}) + 2\hat{\rho}}{T}, \quad (2)$$

where m is the order of a polynomial describing the X variable. In the case of the mean or regressions against an untrended explanatory variable, m was set to zero; for regressions against time m was set to one. Whenever inadmissible estimates were obtained [i.e., outside $(-1, 1)$], $\hat{\rho}$ was set to -0.99 or 0.99 .

For uncorrelated errors, the usual estimate of the mean is the sample average, $\hat{\mu} = \sum_i y_i / T$. Assuming normality, its confidence interval is

$$\hat{\mu} \pm st_{v,(\alpha/2)}, \quad (3)$$

where s is the usual estimated standard error of the mean

$$s = \sqrt{\frac{\sum_{i=1}^T (y_i - \hat{\mu})^2}{T(T-1)}} \quad (4)$$

and for the desired $100(1 - \alpha)$ percent confidence interval, t is obtained from tables of student's t distribution using $v = T - 1$ df. For the regression case, estimates of μ and β and their standard errors are usually obtained using ordinary least squares (e.g., Draper and Smith 1981), and confidence intervals are obtained from

$$\hat{\theta}_j \pm s_{\theta_j} t_{v,(\alpha/2)}, \quad (5)$$

where $\theta = (\theta_1, \theta_2) = (\mu, \beta)$, s_{θ_j} is the standard error for the specified parameter, and $v = T - 2$. I will call these estimators, including the sample mean, ordinary least squares (OLS). All the adjusted methods for constructing confidence intervals described below use Eqs. 3 and 5, differing only in the estimates of the parameters and their standard errors.

The simplest correction method I call corrected OLS, and apply it only to estimating the mean. This method still uses the sample average as an estimate of μ , but multiplies s by a correction factor k , either by adjusting the variance or the effective sample sizes (e.g., Bayly and Hammersley 1946, Stewart-Oaten 1986). A priori, this corrected OLS method seems a reasonable contender since the OLS mean has efficiency near what can be obtained with the minimum variance unbiased estimator (Chipman et al. 1968, Fishman 1972, Mack 1986). Under the usual assumptions, s is an estimate of the standard deviation of \bar{y} . We want k so that $E(ks)$

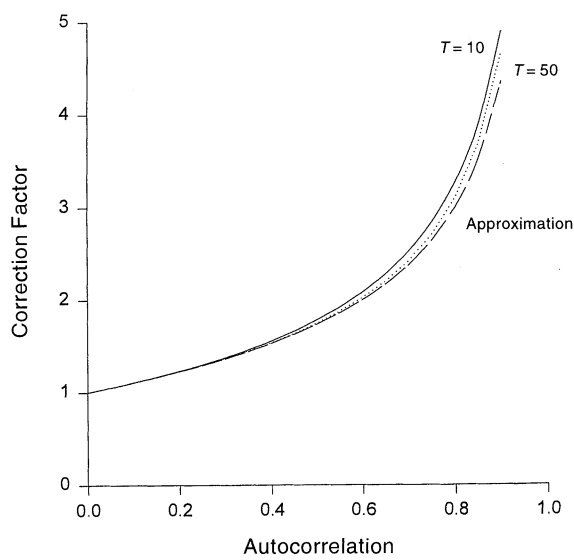


FIG. 1. Correction factors (assuming known autocorrelation) for adjusting the usual standard error of a sample mean. Shown are exact correction factors for sample size $T = 10$ and sample size $T = 50$ and an approximation.

equals the standard deviation of \bar{y} when the $\epsilon_i = y_i - \mu$ come from an AR(1) process. A correction factor that approximates this for moderate-to-large sample sizes is $k = [(1 + \rho)/(1 - \rho)]^{1/2}$, and a correction factor for which this holds exactly is

$$k = \left[\frac{1 + 2\delta/T}{1 - 2\delta/T(T-1)} \right]^{1/2} \tag{6}$$
$$\delta = \frac{[(T-1)\rho - T\rho^2 + \rho^{T+1}]}{(1-\rho)^2}$$

(Stewart-Oaten 1986). Fig. 1 gives the approximate and exact correction factors (for sample sizes of 10 and 50) as a function of ρ . Fig. 1 is also illustrative of just how far from correct s (and thus also confidence intervals) can be if the autocorrelation is ignored. Of course, in virtually all applications, ρ is replaced by an estimate.

A second method of correction is by two-stage estimation (Cochrane and Orcutt 1949). First, ρ is estimated from the OLS residuals, and this estimate is then substituted for ρ , acting as though it were known. Estimation of the parameters θ [θ equals μ or (μ, β)] is by generalized least squares, minimizing over θ

$$S(\theta) = (1 - \hat{\rho}^2)[y_1 - f(\theta, x_1)]^2 + \sum_{i=2}^T \{[y_i - f(\theta, x_i)] - \hat{\rho}[y_{i-1} - f(\theta, x_{i-1})]\}^2, \tag{7}$$

where $f(\theta, x_i)$ is μ or $\mu + \beta x_i$ for estimating the mean or regression parameters, respectively. These estimates and their standard errors can be obtained by doing an ordinary least squares regression on a transformed Y

and X . A clear description of this approach is given by Seber and Wild (1989).

The iterated two-stage estimator is a variant of the above estimator. The estimate of ρ is recalculated based on residuals about the two-stage estimates of the $f(\theta, x_i)$'s, and then an updated estimate of θ is produced by substituting this new estimate of ρ into the generalized least squares procedure. These steps are then iterated until the estimate of ρ converges. This procedure finds the values of θ and ρ that jointly minimize Eq. 7, because the Prais-Winsten estimate of autocorrelation is the value of $\hat{\rho}$ that minimizes Eq. 7 for a given θ .

Estimation of θ and the associated standard error can also be done by maximum likelihood. The log likelihood equation, assuming normality, is

$$L(\theta, \rho, \sigma_\epsilon^2) = \text{const} - \frac{T}{2} \log \sigma_\epsilon^2 + \frac{1}{2} \log(1 - \rho^2) - \frac{1}{(2\sigma_\epsilon^2)} S(\theta, \rho), \tag{8}$$

with $S(\theta, \rho)$ defined as in Eq. 7, now with ρ replacing $\hat{\rho}$ and treated as a parameter, and the maximization is over θ , ρ , and σ_ϵ^2 . I maximized this log likelihood equation using a standard quasi-Newton procedure. (Note that a concentrated likelihood estimation approach can be more efficient for linear models; see the discussion in Seber and Wild [1989: 316–317].) Asymptotically valid estimates of the standard errors for θ were obtained from the Hessian matrix (e.g., Seber and Wild 1989). Note that the difference between the iterated two-stage and maximum likelihood approaches is in the $\log(1 - \rho^2)$ term of Eq. 8. This term arises because the maximum likelihood estimator treats the first observation as coming from its stationary distribution, while in the estimation of ρ the iterated two-stage approach takes the first observation as given. The $\log(1 - \rho^2)$ term enforces stationarity and acts to bound the maximum likelihood estimate away from -1 and $+1$.

Another potential method for producing confidence limits of the appropriate size is the use of computer-intensive resampling procedures such as Monte Carlo simulations, the jackknife or the bootstrap (see Crowley 1992 for a review). Carpenter et al. (1989) suggest estimation of autocorrelation followed by a Monte Carlo approach for ecological time series. Freedman and Peters (1984) discuss the application of the bootstrap in a multiequation time series context, and in their application the resulting confidence intervals were much closer to correct than those obtained from a two-stage estimator. Efron and Tibshirani (1986) illustrate an application of the bootstrap procedure to an autoregressive process. Rayner (1990) showed that the bootstrap worked well for determining P values for hypothesis tests on the autocorrelation coefficient of an AR(1) process, even when sample sizes were quite small. The key feature of applying the (ordinary) bootstrap to a potentially autocorrelated time series is that residuals

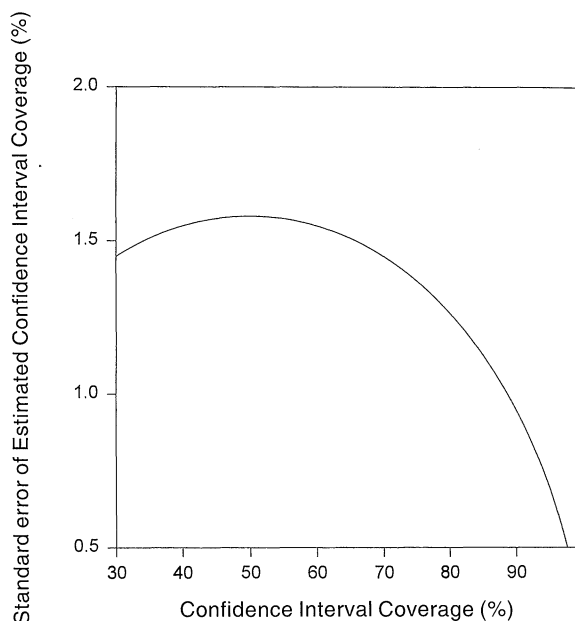


FIG. 2. Standard error of estimated coverage of confidence interval based on 1000 replicates, as a function of the actual coverage of the confidence interval.

from the time series model (which are asymptotically independent and identically distributed) are resampled, not the original autocorrelated observations.

MONTE CARLO EXPERIMENTS

The true size of nominal 95% confidence intervals for the above methods was explored in several experiments. In most experiments T was set at 10, 25, and 50, and for each T , ρ took values of 0, 0.1, 0.3, 0.5, 0.7, and 0.9. One thousand data sets were generated in each case. For this number of replications the standard error of the estimated coverage is 0.7% when the coverage is 95%, and the standard error of the estimated coverage for 1000 replications is plotted as a function of the coverage in Fig. 2. In each case y_1 was generated from its stationary gaussian distribution [i.e., with mean zero and variance $\sigma_a^2/(1 - \rho^2)$]. The y_i , $i > 1$, were generated from the AR(1) process $y_i = \rho y_{i-1} + a_i$, where the a_i are independent gaussian errors with variance σ_a^2 .

All simulations were done in the Gauss statistical simulation language. Where feasible, routines were checked against hand-calculated values. Other quantities (e.g., maximum likelihood and iterated two-stage estimates) were compared with standard output from the SAS autoreg procedure (SAS Institute 1984). To make these comparisons with SAS output for the two-stage and iterated two stage methods, I modified my programs to replace the Prais-Winsten estimate of the autocorrelation with the Yule-Walker estimate used by SAS, which differs from Eq. 1 by including the first and n^{th} terms in the denominator sum. For small sam-

ples this alternative estimator is substantially more biased, and is not recommended (see *Results* and Park and Mitchell [1980]).

The first experiment evaluated the corrected OLS and maximum likelihood methods for estimating the sample mean. A second experiment evaluated the two-stage and iterated two-stage estimators. In practice, often OLS is first fit, and the residuals are tested for significant autocorrelation, with OLS being replaced by a "corrected" method only when significant autocorrelation is detected. Consequently, variants of the methods following this approach were also examined. Autocorrelation was tested for using the Durbin-Watson statistic Q , with P values determined by matching the first two moments of the linear transformation $a + bQ$ to those of a beta distribution (Henshaw 1966, Durbin and Watson 1971). Pretesting for autocorrelation was done at both the conventional level of 0.05 and at 0.50 (Fomby and Guilkey 1978), and when the autocorrelation was declared statistically significant the OLS method was replaced by the specified method.

In a third experiment, bootstrap estimates of confidence intervals for the mean were evaluated using only the two-stage method. One thousand sets of data were generated, and for each data set 1000 bootstrap samples were taken to estimate the confidence interval. Here I used the percentile method to construct confidence intervals, rather than the more involved BC or BC_a methods, because neither bias nor skewness are significant problems for the estimate of the mean (see Efron and Tibshirani 1986). Because of the computationally intensive nature of bootstrapping simulation results (10^6 estimates per case), I only examined the $T = 10$, $\rho = 0.7$ case as an example; this case was chosen because there was serious undercoverage in experiment one.

Results of the first two experiments, and results of other studies (e.g., King 1986, Wilson 1989, Doran et al. 1992, Al-Subhi 1993) suggest that bias in estimation of ρ can cause problems. Therefore, in a fourth experiment, corrected OLS, two-stage, and iterated two-stage estimators of the mean were evaluated using a bias-corrected estimator of ρ . The bias-corrected two-stage estimator was suggested by Doran et al. (1992).

Experiments five through eight examined the coverage of confidence intervals of intercept and slope estimates from a linear regression. In the fifth and sixth experiments, regressions were against "time" as represented by the index i associated with each y_i . This was chosen as a strongly trended explanatory variable. In the fifth experiment I used the same suite of estimators as in experiments one and two (excluding the corrected OLS method), and in the sixth experiment I used the same estimators as in experiment four. Experiments seven and eight were the same as five and six, but now the explanatory variable was untrended "white noise," i.e., independent values from a normal distribution. For all the regression experiments the true

TABLE 1. Estimated coverage (%) of nominal 95% confidence intervals based on 1000 simulation trials, for each combination of sample size (T), pretest level (PT), and autocorrelation (ρ). Methods are: ordinary least squares (OLS), corrected OLS (Corr. OLS), two-stage estimator (2-stage), iterated two-stage estimator (I. 2-stage) and maximum likelihood (ML). PT = N means no pretest was done.

Method	Pre-test	Autocorrelation (ρ)																	
		$T = 10$						$T = 25$						$T = 50$					
		0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9
Corr OLS	N	90	90	87	84	77	56	95	93	90	90	83	73	96	96	94	92	90	83
2-stage	N	91	91	88	84	74	52	95	93	91	85	83	70	95	94	94	92	89	76
I. 2-stage	N	91	91	88	84	75	53	95	93	91	85	83	71	95	94	94	92	89	77
ML	N	89	89	86	83	74	54	94	92	91	89	83	72	95	95	96	93	89	83
Corr OLS	0.5	95	94	91	85	77	57	97	94	91	90	91	73	97	96	94	92	91	83
2-stage	0.5	95	95	90	85	74	52	97	95	91	89	83	70	96	94	95	92	89	76
I. 2-stage	0.5	95	95	90	85	75	53	97	95	91	89	83	71	96	94	96	92	89	77
ML	0.5	95	94	89	84	75	55	96	94	91	88	90	72	96	96	96	93	89	83
Corr OLS	0.05	95	95	88	82	72	55	94	89	90	83	83	73	96	96	94	92	90	83
2-stage	0.05	95	96	90	85	74	52	94	89	91	87	83	70	96	96	91	92	89	76
I. 2-stage	0.05	95	96	90	85	75	53	94	90	91	87	83	71	96	96	91	92	89	77
ML	0.05	95	95	88	88	72	53	94	89	88	82	82	77	96	96	95	92	89	83

slope was zero, i.e., Y values were generated with no underlying relationship to the explanatory X .

RESULTS

Confidence intervals for the mean

As expected, the estimated coverage (i.e., the percent of the time the confidence interval contained the true value) of the OLS confidence intervals of the mean was near 95% for $\rho = 0$, and decreased as ρ increased, to values near 30% for $\rho = 0.9$ and sample sizes of 10 or 25 (Table 1). Of more interest is the coverage of the methods that “correct” for autocorrelation. In the presence of moderate-to-high autocorrelation, all these methods used in experiments 1 and 2 produced confidence intervals with substantially less than the 95% nominal coverage (Table 1). Even for sample sizes of 50, coverage fell below 90% for $\rho = 0.9$. Pretesting led to true coverage nearer 95% for small sample sizes

and mild autocorrelation and worked best at the 0.50 level, but the improvement was only modest. The undercoverage was quite similar for all estimators but OLS, and in all but a few cases the small differences (excluding OLS results) can be reasonably attributed to chance (compare Fig. 2 and Table 1).

Part of the cause for undercoverage is that all the estimates of ρ are biased downward, with the bias being stronger for smaller sample sizes and stronger autocorrelation (Table 2). Thus we are tending to “under-correct.” The situation is more complex than simply underestimating ρ , since the estimates vary about their expected value, and for positive autocorrelation are skewed to the left (e.g., Fuller 1976, Fig. 3a). To illustrate how the undercorrection comes about, I present results from the corrected OLS estimator for one case ($\rho = 0.7, T = 25$) in greater detail. Fig. 3a gives the distribution of estimated autocorrelations. A majority of estimates are less than the true value, and there is a substantial skew in the distribution.

These two properties of the distribution contribute to the estimated bias. The underestimates of ρ lead to undercorrection (Fig. 3b), so the majority of estimated “correction factors” are less than the “true” correction factor [i.e., the value k so that $E(ks) = \sigma_\mu$]. Note that the distribution of correction factors is skewed to the right. This reversal in skew occurs because k is an accelerating function of ρ (Fig. 2). The assumption under which the confidence intervals are constructed is that $(\hat{\mu} - \mu)/\hat{\sigma}_\mu = \hat{\mu}/\hat{k}s$ follows a t distribution on $T - 1$ df. The effect of the frequent undercorrections is to make the tails of the distribution of $\hat{\mu}/\hat{k}s$ heavier than the assumed t distribution (Fig. 3c), and this is what leads to the undercoverage of the confidence intervals. In this example, large overcorrections are not common enough to produce a noticeable increase in the frequency near zero.

TABLE 2. Means of 1000 estimates of ρ (with no pretesting) associated with different methods of estimating μ , for each combination of sample size (T), and autocorrelation (ρ). Methods are: $\hat{\rho}$ based on ordinary least squares residuals (OLS), $\hat{\rho}$ based on iterated two-stage residuals (I. 2-stage), and $\hat{\rho}$ estimated by maximum likelihood (ML). Note that the corrected OLS and two-stage method use the OLS estimate of ρ .

T	Method	Autocorrelation (ρ)					
		0	0.1	0.3	0.5	0.7	0.9
10	OLS	-0.11	-0.04	0.09	0.24	0.37	0.48
	I. 2-stage	-0.12	-0.05	0.10	0.24	0.47	0.48
	ML	-0.11	-0.04	0.09	0.25	0.38	0.50
25	OLS	-0.02	0.05	0.20	0.39	0.56	0.73
	I. 2-stage	-0.03	0.07	0.23	0.38	0.58	0.74
	ML	-0.03	0.05	0.20	0.39	0.56	0.75
50	OLS	-0.02	0.06	0.24	0.43	0.64	0.82
	I. 2-stage	-0.01	0.07	0.27	0.45	0.64	0.82
	ML	-0.02	0.08	0.27	0.44	0.64	0.82

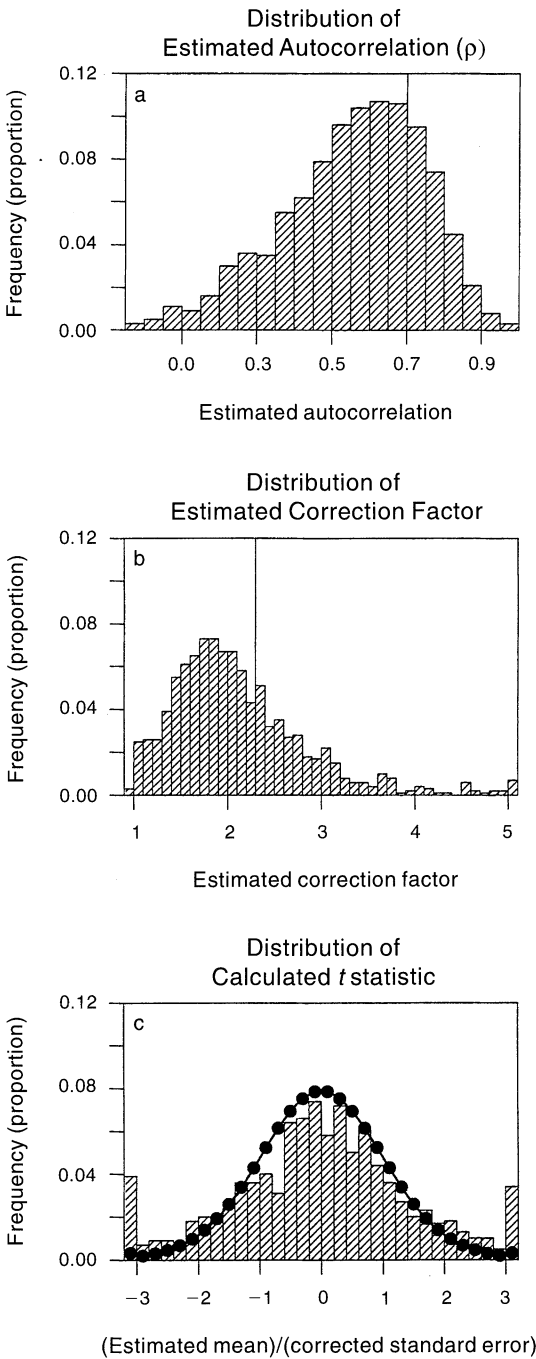


FIG. 3. Estimated frequency distributions for the corrected OLS method for $\rho = 0.7$ and $T = 25$. The X axis represents the midpoint of intervals. In each case the first and last intervals are accumulator bins containing all values less than or more than the indicated bin, respectively. (a) Frequency distribution of estimated values of the autocorrelation (ρ). Vertical line represents the true value. (b) Frequency distribution of estimated values of k , the correction factor. Vertical line indicates the “correct” value given by Eq. 6 for $\rho = 0.7$. (c) Frequency distribution of t statistics $\hat{\rho}/k/s$. Circles connected by solid line indicate the predicted frequencies from a t distribution with $T - 1 = 24$ df.

TABLE 3. Estimated coverage (%) of nominal 95% confidence intervals using bias-corrected methods based on 1000 simulation trials, for each combination of sample size (T), and autocorrelation (ρ). Shown are results for a bias-corrected variant of corrected OLS estimator (B. C. OLS), the bias-corrected two-stage estimator (B. 2-stage), and the bias-corrected iterated two-stage estimator (BI. 2-stage).

T	Method	Autocorrelation (ρ)					
		0	0.1	0.3	0.5	0.7	0.8
10	B. C. OLS	92	90	91	87	81	82
	B. 2-stage	97	96	93	89	85	73
	BI. 2-stage	97	96	93	89	87	77
25	B. C. OLS	94	94	91	91	91	95
	B. 2-stage	96	96	94	93	90	83
	BI. 2-stage	96	96	94	93	90	85
50	B. C. OLS	93	94	93	93	95	99
	B. 2-stage	96	96	93	93	93	87
	BI. 2-stage	96	96	93	93	93	88

In a third experiment a bootstrap estimator of the confidence interval was applied only to the two-stage method for $\rho = 0.7$ and $T = 10$, with a pretest at the 0.5 level. The percentage of the bootstrap confidence limits that overlapped zero (the true value of μ) was 52.6%, which was substantially lower than the coverage of the confidence intervals for the conventional two-stage method (Table 1). The bootstrap procedure used a conventional estimate of the autocorrelation, and thus was not different than conventional approaches with respect to bias of autocorrelation estimates. An additional problem for the bootstrap is that for small samples the variability of the resampled residuals about the fitted model is less than the variability of the actual random errors.

In experiment 4 I evaluated methods using a bias-corrected estimate of the autocorrelation. As in experiments 1 and 2, each method used in experiment 4 was done without a pretest, and with pretests at the 0.05 and 0.5 levels. Results with and without pretests were similar, with slightly better coverage for small sample sizes and modest autocorrelation when a pretest was used. For the brevity I report here (Table 3) the results of the estimators with a 0.50 pretest.

The use of the bias correction was effective in bringing the actual coverage of the confidence intervals for the mean closer to the nominal level for each method (Table 3), but none of the methods completely solved the problem of undercoverage for sample sizes of 25 or less. At times, the procedure can be awkward to apply; for short time series and strong autocorrelation there is a reasonable probability that the bias-corrected estimates of ρ will cross the 1.0 bound. The arbitrary value (0.99) these estimates were set to in the simulations greatly influences the size of the confidence intervals. It might be best to conclude only that the confidence bounds are wide but not well estimated in such cases. As a practical matter, for most short time series it might not matter much whether ρ is set to 0.99

TABLE 4. Estimated coverage (%) of nominal 95% confidence intervals for intercept (μ) and slope (β) parameters (Par.) of regressions vs. time based on 1000 simulation trials, for each combination of sample size (T), and autocorrelation (ρ). Methods are: iterated two-stage estimator (I. 2-stage), bias-corrected two-stage estimator (B. 2-stage), bias-corrected iterated two-stage estimator (BI. 2-stage), and bias-corrected iterated two-stage estimator with a pretest at the 0.5 level (PBI. 2-stage).

Method	Par.	Autocorrelation (ρ)																			
		$T = 10$						$T = 25$						$T = 50$							
		0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9		
I. 2-stage	μ	87	87	81	77	67	49	93	92	88	89	79	60	92	93	91	90	86	69		
B. 2-stage	μ	89	88	87	82	75	69	93	93	93	90	86	76	94	95	94	93	91	85		
BI. 2-stage	μ	89	89	88	85	79	77	93	93	93	90	87	82	94	95	94	93	92	87		
PBI. 2-stage	μ	97	94	92	88	81	77	95	95	94	91	87	82	95	95	94	93	92	87		
I. 2-stage	β	87	87	81	78	73	58	92	92	89	84	81	64	92	93	91	90	86	69		
B. 2-stage	β	90	89	88	83	78	69	93	93	92	91	86	79	94	95	94	93	92	85		
BI. 2-stage	β	90	89	88	85	81	77	93	93	92	92	88	84	94	95	94	93	92	88		
PBI. 2-stage	β	96	95	92	88	82	77	96	95	93	92	88	84	96	96	94	93	92	88		

or 0.9999, because in either case the confidence bounds will be so broad that little could be claimed on the basis of the estimate.

No one bias-corrected estimator performed best in all situations (Table 3). The bias-corrected variants of the two stage and iterated two-stage estimators performed similarly and, for sample sizes of ≤ 25 and moderate autocorrelation, there was a tendency for them to outperform the bias-corrected variant of corrected OLS. This difference, however, was not large. For strong autocorrelation (0.9), the bias-corrected variant of corrected OLS produced higher coverage than the other methods, and for a sample size of 50 actually had coverage substantially above 95%.

Confidence intervals for regression parameters

Park and Mitchell (1980) previously evaluated the two-stage, iterated two-stage, and maximum likelihood estimators for the regression model, including regression against time. They found that the iterated two-stage estimator outperformed the others to a modest degree, a result that my simulations agree with. Con-

sequently, among these methods results are presented only for the iterated two-stage estimator. Results are also presented for the bias-corrected two-stage and bias-corrected iterated two-stage estimators. These results are presented without pretests to facilitate comparisons with results of Park and Mitchell (1980) and Doran et al. (1992). Results are also presented for the bias-corrected iterated two-stage estimator with a 0.50 pretest since this was the best method overall for matching claimed coverage of the confidence intervals.

In regressions against time, even for the best method, for $T = 10$ the actual coverage of confidence intervals of the intercept fell below 85% for $\rho \geq 0.7$ (Table 4). For $T = 25$ the coverage fell below 85% at $\rho = 0.9$, and for $T = 50$ coverage fell below 90% when $\rho = 0.9$. The results were similar for estimates of the slope (Table 4). Modest improvement was obtained by pretesting.

For regressions against untrended white noise, the results for the intercept were much the same as for the regression against time, and again there was substantial undercoverage (Table 5). Again the bias-corrected iterated two-stage estimator with a pretest at the 0.5 level

TABLE 5. Estimated coverage (%) of nominal 95% confidence intervals for intercept (μ) and slope (β) parameters (Par.) of regressions vs. untrended "noise" based on 1000 simulation trials, for each combination of sample size (T), and autocorrelation (ρ). Methods are: iterated two-stage estimator (I. 2-stage), bias-corrected two-stage estimator (B. 2-stage), bias-corrected iterated two-stage estimator (BI. 2-stage), and bias-corrected iterated two-stage estimator with a pretest at the 0.5 level (PBI. 2-stage).

Method	Par.	Autocorrelation (ρ)																			
		$T = 10$						$T = 25$						$T = 50$							
		0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9	0	0.1	0.3	0.5	0.7	0.9		
I. 2-stage	μ	89	91	84	85	76	56	93	91	92	90	86	63	94	94	93	92	89	78		
B. 2-stage	μ	91	92	93	86	79	67	93	94	91	93	86	78	93	94	95	92	90	85		
BI. 2-stage	μ	91	93	94	88	84	76	93	94	91	93	89	84	93	94	95	93	92	88		
PBI. 2-stage	μ	96	96	95	89	85	77	95	95	92	93	89	84	96	95	95	93	92	88		
I. 2-stage	β	90	90	91	94	96	97	93	94	92	95	95	96	94	94	95	95	96	95		
B. 2-stage	β	91	92	90	95	95	94	94	94	93	95	95	95	94	95	94	96	95	94		
BI. 2-stage	β	89	92	89	94	94	94	94	93	93	95	95	94	94	95	94	96	95	94		
PBI. 2-stage	β	91	92	91	95	95	94	95	94	93	95	95	94	95	95	94	96	94	94		

was best at matching nominal coverage. In contrast, the actual coverage of confidence intervals of the slope were generally close to the stated coverage for all the estimators shown in Table 5. The only problem appears to be a moderate amount of undercoverage for $T = 10$ when the autocorrelation was low.

Efficiency

In general, the efficiency of parameter estimates, as well as the true coverage of their nominal confidence intervals, should play a role in choosing among estimators. Relative efficiency of the estimators considered here, as measured by the ratio of the root mean square error (RMSE) of the estimate to the RMSE for the OLS estimate, did not differ substantially from 1.0 in most cases. The one exception was for slope estimates in regressions against trendless white noise. Here, the relative efficiency of both the conventional and bias-corrected iterated two-stage estimators increased substantially above 1.0 for sample sizes >25 and autocorrelation stronger than 0.7 (Table 6).

ANALYSIS OF EXAMPLE DATA SETS

Two contrasting examples were selected. The first example (Fig. 4a) is for the case of estimating the mean, and comes from baseline sampling of large benthic invertebrates living in the shallow rocky subtidal zone offshore from the San Onofre Nuclear Generating Station in southern California (Schroeter et al. 1993). The data are on the relative abundance of the white sea urchin (*Lytechinus anamesus*), and come from nine surveys that were done at ≈ 3 -mo intervals prior to the start of operations by the generating station's two new units in mid-1983. Each data point represents the difference in log-transformed density (numbers per unit area) between a Control site (Barn kelp forest ≈ 11 km from the station) and an Impact site (the upcoast portion of San Onofre kelp forest). The idea was to estimate the mean difference and its standard error or confidence interval prior to any influence of the new units. This

TABLE 6. Efficiency relative to OLS (ordinary least squares) estimator of two candidate estimators of the slope (β) of a linear regression against trendless "noise" for each combination of sample size (T) and autocorrelation (ρ). Relative efficiency is defined by the ratio of the root mean square error (RMSE) of the OLS estimator to the RMSE of the candidate estimator. The candidate estimators are the iterated two-stage estimator (I, 2-stage) and the bias-corrected iterated two-stage estimator (BI, 2-stage).

T	Method	Autocorrelation (ρ)					
		0	0.1	0.3	0.5	0.7	0.9
10	I, 2-stage	0.98	0.92	0.99	1.11	1.15	1.21
	BI, 2-stage	0.89	0.92	0.98	1.12	1.19	1.62
25	I, 2-stage	0.98	0.98	1.08	1.20	1.74	3.18
	BI, 2-stage	0.95	0.98	1.04	1.14	1.49	1.82
50	I, 2-stage	1.00	0.99	1.08	1.27	1.65	3.45
	BI, 2-stage	0.98	0.98	1.04	1.29	1.53	2.16

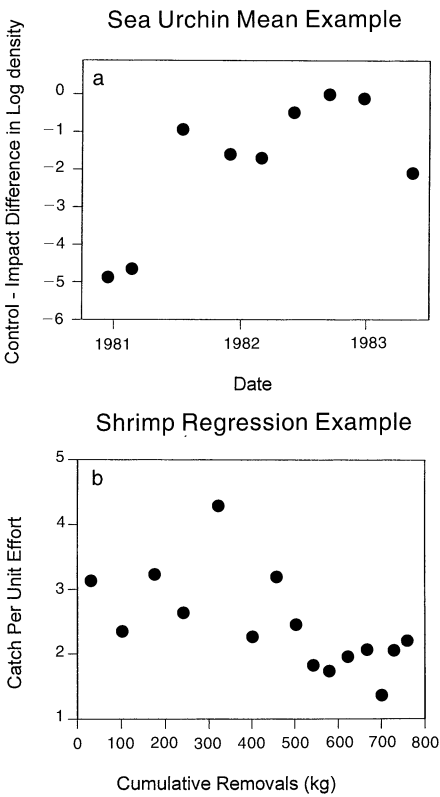


FIG. 4. Two example data sets. (a) The difference in log-transformed density (numbers per square metre) of white sea urchins between an impact site (upcoast San Onofre kelp forest) and a potential control site (Barn kelp forest) for nine surveys (Schroeter et al. 1993). (b) The relationship between catch per unit effort (catch biomass per trap) and cumulative biomass (kilograms) of removals from a population of caridean shrimp (Ralston 1986). As a continuity correction, cumulative removals on day i are defined as including half the removals on the i^{th} day (see Ralston 1986).

mean difference then could serve as a basis for comparison with differences observed after the new units began operating. The underlying rationale for this approach is described in the context of a time series process by Stewart-Oaten et al. (1986).

An important lesson emphasized by the analysis of this data set is that although it is important to adjust for autocorrelation, such adjustments can be difficult and problematic when time series are quite short and autocorrelation is strong.

For this example, a reasonable procedure based on the simulations is to use the bias-corrected two-stage estimator after a pretest at the 0.5 level. In this case there was strong evidence for positive autocorrelation. The estimated autocorrelation (ρ) based on the OLS residuals was 0.70 and the P value for the Durbin-Watson test was 0.0001. The final estimate of autocorrelation after bias correction exceeded 1.0 (1.04) and was set to 0.99. Table 7 gives the estimated means and 95% confidence intervals from the bias-corrected

TABLE 7. Parameter estimates and nominal 95% confidence intervals for two example data sets. Methods are: ordinary least squares (OLS), two-stage (2-stage), bias-corrected two stage (B. 2-stage), and the bias-corrected variant of the iterated two-stage estimator (BI. 2-stage).

Parameter	Method	Confidence interval
Sea urchin mean example (Schroeter et al. 1993)		
μ	OLS	-1.83 ± 1.36
	2-stage	-2.40 ± 3.25
	B. 2-stage	-3.41 ± 26.26
Shrimp regression example (Ralston 1986)		
μ	OLS	3.33 ± 0.77
	2-stage	3.35 ± 0.65
	BI. 2-stage	3.35 ± 0.66
β	OLS	-0.0019 ± 0.0015
	2-stage	-0.0020 ± 0.0012
	BI. 2-stage	-0.0020 ± 0.0012

two-stage estimator (using $\rho = 0.99$) and, for comparison, from the widely used OLS and conventional two-stage estimators. The results show that the choice of estimation method strongly influences both the point estimate and the size of the confidence interval, with the bias-corrected variant of the two-stage method having a much larger confidence interval than the other methods.

As part of a simulation procedure, setting estimates of ρ exceeding 1.0 to an arbitrary value like 0.99 seems reasonable. This becomes more awkward during the analysis of a real data set. Quite different confidence intervals would be obtained if different arbitrary values, say 0.999 or 0.9999, were used instead. However, the simulation results show us that falling back on the conventional two-stage estimator is likely to lead to greater undercoverage. In spite of the awkwardness in this application, the bias correction results lead to the conclusion that the confidence interval is poorly established, and potentially should be much broader than is suggested by the conventional estimators.

The second example data set comes from intensive experimental removals from a small isolated population of caridean shrimp in the Mariana Archipelago (Ralston 1986). Removals were done by trapping over a 15-d period, and data on catch biomass per unit effort (CPUE), i.e., per trap, and total removals were collected each day (Fig. 4b). The aim of the experiment was to estimate catchability of the gear being used, and to estimate total population biomass. The analysis follows the Leslie method (see Ricker 1975). This method is based on the assumption that expected $CPUE_i = qB_i = q(B_0 - R_i)$, where q is catchability, B_i is the biomass at time i , B_0 is the initial biomass, and R_i are the removals up to time i . For the estimated regression equation $E(CPUE_i) = \hat{\mu} + \hat{\beta}R_i$, estimates of q and B_0 are given by $-\hat{\beta}$, and $-\hat{\mu}/\hat{\beta}$, respectively. In this experiment, the cumulative removals (R_i) increased approximately linearly over time, and thus the explanatory variable is strongly trended. A value of $m = 1$ was

used in the bias correction methods. In contrast with the first example, there is no evidence for positive autocorrelation in this case. The estimated autocorrelation was -0.19 for OLS, and -0.17 for the bias-corrected iterated two-stage approach. The Durbin Watson P value was 0.64. It is reasonable to use OLS here, because the pretest was not significant at the 0.5 level. If a correction method were used, the bias-corrected variant of the iterated two-stage estimator would be the choice because of its relatively good performance in the simulations. These estimates and confidence intervals, along with a conventional two-stage estimate for comparison, were all similar (Table 7). In spite of this lack of sensitivity to estimation method, the simulation results suggest that the confidence intervals need to be interpreted cautiously.

DISCUSSION

Ignoring positively autocorrelated errors and using the usual confidence intervals (or hypothesis tests) can lead to serious undercoverage, or inflated Type I error rates. This is well known, although as Hurlbert (1984) pointed out, it is a problem that is often ignored. Perhaps more disturbing is that reasonable methods to adjust confidence intervals and tests often undercorrect for the problem, especially when sample sizes are small. The problem appears, in part, to be related to bias in the estimation of the autocorrelation coefficient. This general result is well known in the econometric literature, but is not as well known to ecologists.

Of the methods evaluated here, approaches using bias-corrected estimates of autocorrelation had true coverage closest to the nominal 95%. When autocorrelation was strong, the best methods still displayed substantial undercoverage, even for sample sizes of 25–50. The only exception was for estimates of the slope when the explanatory variable was untrended white noise. Doran et al. (1992) provide a theoretical explanation for this exception.

For very small sample sizes (on the order of 10) and strong autocorrelation, the bias correction approach can be awkward because corrected estimates of autocorrelation will frequently be greater than 1.0. In such cases, the reality is that not only is there great uncertainty associated with the parameter estimates but also with their standard errors and confidence intervals.

As noted above, concern about undercorrection by standard “corrected” methods has been repeatedly expressed in the statistical and econometric literature. When Cochran and Orcutt (1949) introduced a variant of the two-stage estimator, they pointed out that bias in estimation of the autocorrelation coefficient limited the applicability of the method; they in fact speculated on the use of a bias correction method. For a number of years, the warning of Cochran and Orcutt was largely ignored, as the two-stage estimator became a standard method for fitting regressions to time series. Since then, many papers have discussed the bias of estimated

autocorrelation and speculated on the implications for confidence intervals and hypothesis tests (e.g., Kendall 1954, Marriot and Pope 1954, Kobayashi 1985, Shaman and Stine 1988). There have also been a number of direct examinations of the validity of confidence intervals for the standard methods used to analyze autocorrelated data, mostly in a regression context (e.g., Nakamura and Nakamura 1978, Park and Mitchell 1980, Griffiths and Beesley 1984, King and Giles 1984, Doran et al. 1992, Al-Subhi 1993).

The results of this study are not the final word on the best way to deal with autocorrelation, and this remains an active area of research. As pointed out by Fomby and Guilkey (1978), discussed by Doran et al. (1992), and illustrated here, the performance of an estimator can depend upon the type of explanatory variable (e.g., untrended, trended, or correlated in time). Furthermore, there are many other ways to construct confidence intervals in addition to those used here.

Two promising (and related) alternative approaches are conditional marginal likelihood (CML) (e.g., Wilson 1989) and Bayesian methods (e.g., Severini 1993). For CML, the autocorrelation coefficient with the maximum of the marginal likelihood is used as an estimate of ρ in a two-stage approach. Marginal likelihood has been advocated as a means for treating nuisance parameters and reducing bias in parameters of interest (Kalbfleisch and Sprott 1970). Doran et al. (1992) suggested using the mean of the Bayesian posterior distribution as an estimate of the autocorrelation in a two-stage approach. This is related to CML because the Bayesian posterior distribution is proportional to the marginal likelihood (e.g., Levenbach 1972). A more fully Bayesian approach is to determine the Bayesian posterior density of the parameter of interest (e.g., the mean or regression slope), and treat the $100(1 - \alpha)\%$ posterior density interval as though it is a confidence interval.

Al-Subhi (1993) examined CML and this "fully" Bayesian approach along with three other methods, only one of which (maximum likelihood) overlaps with the methods used here. He examined the performance of the methods for estimating the mean, the slope of a regression against time (t), and in a quadratic regression, the slope against t^2 . Of the methods he considered, conditional marginal likelihood (CML) and the Bayesian method worked best in producing correct coverage. He recommended the conditional marginal likelihood method because it was computationally less intensive.

Comparison of the results in this paper with those of Al-Subhi (1993) for overlapping sample sizes and autocorrelation indicates that the CML method does not have as good coverage properties as the best procedures examined here for estimating the mean. In contrast, for estimating the slope of a regression the CML procedure appears generally superior. Al-Subhi does not report results for the intercepts of his regressions.

The CML method appears a promising approach, and further work and comparisons are needed. It has an advantage over the bias correction approach in that the order of a polynomial does not need to be specified for the explanatory variable.

Although Al-Subhi's (1993) results show that the Bayesian estimator provides little benefit in terms of confidence interval coverage over CML for the situations he examined, it remains a viable approach. It has the advantage of avoiding autocorrelation estimates that end up at the bound $+1$. In cases where prior information on the likely strength of autocorrelation is available or can be assumed, the Bayesian approach would likely prove an even stronger contender.

In conclusion, confidence intervals estimated for the mean or regression parameters from a short time series should be regarded as only approximate, even in the best case. The results shown here and in other studies suggest that errors will often be in the direction of undercoverage. In all the simulations I assumed (correctly) that the real model was an autoregressive process of order one, and thus the undercoverage of the confidence intervals was not related to identification of the error model. Most practitioners recognize that actual correlation structures will often be more complex, and use the first-order corrections as approximations. The effect of this approximation on confidence interval coverage was not evaluated here.

ACKNOWLEDGMENTS

I appreciate comments by Bill Lenarz, Alec MacCall, Robert Kope, Allan Stewart-Oaten, and three anonymous reviewers on earlier drafts of this paper, and input from the ad hoc quantitative group at the Tiburon Laboratory. Steve Ralston stimulated me to finish the work. Craig Osenberg and Gareth Penn supplied me with hard-to-obtain references and articles, and Steve Ralston and Steve Schroeter suggested and supplied example data sets. Julie Detwiler helped with formatting. This research was supported in part by the Minerals Management Service, U.S. Department of the Interior, under MMS Agreement Number 14-35-0001-30471, the Coastal Toxicology Component of the UC Toxic Substances Research and Teaching Program, and the Marine Review Committee, and the final manuscript was completed with support from the Fisheries Division of the Michigan Department of Natural Resources. This paper does not, however, necessarily express the position of the agencies that provided support.

LITERATURE CITED

- Al-Subhi, M. A. 1993. Estimation and prediction for time series regression models: a comparison of inference procedures. Dissertation. Department of Mathematics, Lancaster University, UK.
- Bayly, G. V., and J. M. Hammersley. 1946. The effective number of independent observations in an autocorrelated series. *Journal of the Royal Statistical Society*, **B**, 8:184-197.
- Bence, J. R., A. Stewart-Oaten, and S. C. Schroeter. *In press*. Estimating the size of an effect from a Before-After-Control-Impact-Pairs design: the predictive approach applied to a power plant study. *In* R. J. Schmitt and C. W. Osenberg, editors. *Ecological impact assessment: conceptual issues and application in coastal marine habitats*. University of California Press, Berkeley, California, USA.
- Botsford, L. W., D. A. Armstrong, and J. M. Shenker. 1989.

- Oceanographic influences on the dynamics of commercially fished populations. Pages 511–565 in M. R. Landry and B. M. Hickey, editors. Coastal oceanography of Washington and Oregon. Elsevier Science, Amsterdam, The Netherlands.
- California Coastal Commission. 1991. Adopted Coastal Commission resolution to further condition permit number 183-73: San Onofre nuclear generating station units 2 and 3. Action date 16 July 1991. California Coastal Commission, 45 Fremont Street, Suite 2000, San Francisco, California, USA.
- Carpenter, S. R. 1989. Replication and treatment strength in whole lake experiments. *Ecology* **70**:453–463.
- . 1990. Large-scale perturbations: opportunity for innovation. *Ecology* **71**:2038–2043.
- Carpenter, S. R., T. M. Frost, D. M. Heisey, and T. K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology* **70**:1142–1152.
- Chipman, J. S., K. R. Kadiyala, A. Madansky, and J. W. Pratt. 1968. Efficiency of the sample mean when residuals follow a first-order stationary Markov process. *Journal of the American Statistical Association* **63**:1237–1246.
- Cochrane, D., and G. H. Orcutt. 1949. Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* **44**:32–61.
- Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecology and Systematics* **23**:405–447.
- Cury, P., and C. Roy. 1989. Optimal environmental window and pelagic fish recruitment success in upwelling areas. *Canadian Journal of Fisheries and Aquatic Sciences* **46**:670–680.
- Dickey, D. A., D. P. Hasza, and W. A. Fuller. 1984. Testing for unit roots in seasonal time series. *Journal of the American Statistical Association* **79**:355–367.
- Doran, H. E., W. E. Griffiths, and P. A. Beesley. 1992. Further results on interval estimation in an AR(1) model. Pages 175–194 in W. Griffiths, H. Lutkepohl, and M. E. Bock, editors. Readings in econometric theory and practice. Elsevier Science, Amsterdam, The Netherlands.
- Draper, N. R., and H. Smith. 1981. Applied regression analysis. John Wiley & Sons, New York, New York, USA.
- Durbin, J., and G. S. Watson. 1971. Testing for serial correlation in least squares regression. III. *Biometrika* **58**:1–19.
- Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**:54–77.
- EPA. 1993. Environmental impact statement (EIS) for designation of a deep water ocean dredged material disposal site off San Francisco, California. U.S. Environmental Protection Agency, Region IX, 75 Hawthorne Street, San Francisco, California, USA.
- Fishman, G. S. 1972. Estimating the mean of observations from a wide-sense stationary autoregressive sequence. *Journal of the American Statistical Association* **67**:402–406.
- Fomby, T. B., and D. K. Guilkey. 1978. On choosing the optimal level of significance of the Durbin-Watson test and the Bayesian alternative. *Journal of Econometrics* **8**:203–213.
- Freedman, D. A., and S. C. Peters. 1984. Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association* **79**:97–106.
- Fuller, W. A. 1976. The statistical analysis of time series. John Wiley, New York, New York, USA.
- Gerodette, T. 1987. A power analysis for detecting trends. *Ecology* **68**:1364–1372.
- Griffiths, W. E., and P. A. Beesley. 1984. The small-sample properties of some preliminary test estimators in a linear model with autocorrelated errors. *Journal of Econometrics* **25**:49–61.
- Henshaw, R. C. 1966. Testing single-equation least squares regression models for autocorrelated disturbances. *Econometrica* **34**:646–660.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187–211.
- Jasby, A. D., and T. M. Powell. 1990. Detecting changes in ecological time series. *Ecology* **71**:2044–2052.
- Kalbfleisch, J. D., and D. A. Sprott. 1970. Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society, Series B*, **32**:175–194.
- Kendall, M. G. 1954. Note on bias in the estimation of autocorrelation. *Biometrika* **41**:403–404.
- King, M. L. 1986. Testing for autocorrelation in linear regression models: a survey. Pages 19–73 in M. L. King and D. E. A. Giles, editors. Specification analysis in the linear model. Rutledge and Kegan, London, England.
- King, M. L., and D. E. A. Giles. 1984. Autocorrelation pre-testing in the linear model: estimation, testing and prediction. *Journal of Econometrics* **25**:35–48.
- Kobayashi, M. 1985. Comparison of efficiencies of several estimators for linear regressions with autocorrelated errors. *Journal of the American Statistical Association* **80**:951–953.
- Levenbach, H. 1972. Estimation of autoregressive parameters from a marginal likelihood function. *Biometrika* **59**:61–71.
- Link, W. A., and J. S. Hatfield. 1990. Power calculations and model selection for trend analysis: a comment. *Ecology* **71**:1217–1220.
- Mack, M. E. 1986. Efficiencies of weighted averages in stationary autoregressive processes. *Journal of the American Statistical Association* **81**:730–735.
- Marriot, F. H. C., and J. A. Pope. 1954. Bias in estimation of autocorrelations. *Biometrika* **41**:390–402.
- Mearns, A. J., M. J. Allen, M. D. Moore, and M. J. Sherwood. 1980. Distribution, abundance, and recruitment of soft-bottom rockfishes (Scorpaenidae: *Sebastes*) on the southern California mainland shelf. *California Cooperative Fisheries Investigations Reports* **21**:180–190.
- Millard, S. P., J. R. Yearsley, and D. P. Lettenmaier. 1985. Space-time correlation and its effect on methods for detecting aquatic change. *Canadian Journal of Fisheries and Aquatic Sciences* **42**:1391–1400.
- Nakamura, A., and M. Nakamura. 1978. On the impact of the tests for serial correlation upon the test of significance for the regression coefficient. *Journal of Econometrics* **7**:199–210.
- Osenberg, C. O., R. J. Schmitt, S. J. Holbrook, K. Abu-Saba, and A. R. Flegal. 1994. Detection of environmental impacts: natural variability, effect size and power analysis. *Ecological Applications* **4**:16–30.
- Park, R. E., and B. M. Mitchell. 1980. Estimating the autocorrelated error model with trended data. *Journal of Econometrics* **13**:185–201.
- Ralston, S. 1986. An intensive fishing experiment for the caridean shrimp, *Heterocarpus laevigatus*, at Alamagan Island in the Mariana Archipelago. *Fishery Bulletin* **84**:927–934.
- Rayner, R. K. 1990. Bootstrapping p values and power in the first-order autoregression: a Monte Carlo investigation. *Journal of Business and Economic Statistics* **8**:251–263.
- Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. *Bulletin* 191. De-

- partment of Environment, Fisheries and Marine Service, Ottawa, Ontario, Canada.
- SAS Institute. 1984. SAS/ETS user's guide. Version 5 edition. SAS Institute, Cary, North Carolina, USA.
- Schroeter, S. C., J. D. Dixon, J. Kastendiek, R. O. Smith, and J. R. Bence. 1993. Detecting the ecological effects of environmental impacts: a case study of kelp forest invertebrates. *Ecological Applications* **3**:331–350.
- Seber, G. A. F., and C. J. Wild. 1989. Nonlinear regression. John Wiley & Sons, New York, New York, USA.
- Severini, T. A. 1993. Bayesian interval estimates which are also confidence intervals. *Journal of the Royal Statistical Society, Series B*, **55**:533–540.
- Shaman, P., and R. A. Stine. 1988. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association* **83**:842–848.
- Stewart-Oaten, A. 1986. The before-after/control-impact pairs design for environmental impact assessment. Technical Report prepared for the Marine Review Committee. California Coastal Commission, 45 Fremont Street, San Francisco, California, USA.
- . 1987. Assessing effects on fluctuating populations: tests and diagnostics. In *ASA/EPA conferences on interpretation of environmental data. III. Sampling and site selection in environmental studies* (14–15 May 1987). U.S. Environmental Protection Agency Office of Policy, Planning and Evaluation Publication **EPA-230-08-88-035**.
- Stewart-Oaten, A., W. W. Murdoch, and K. E. Parker. 1986. Environmental impact assessment: “pseudoreplication” in time? *Ecology* **67**:929–940.
- Wilson, G. T. 1989. On the use of marginal likelihood in time series model estimation. *Journal of the Royal Statistical Society, Series B*, **51**:15–27.